10.20378/irb-117467

LLM meets BSLVC: Does *ChatGPT* share grammatical preferences with selected varieties of English?

Katharina Deckert & Fabian Vetter
University of Bamberg, Germany

Relevance

"Generative AI tools will readily act upon users' requests to differentiate between language varieties [...]." (Alvero et al. 2025: 58)

Research question

Can *ChatGPT* match human judgments about the frequency of grammatical features across English varieties?

Previous research

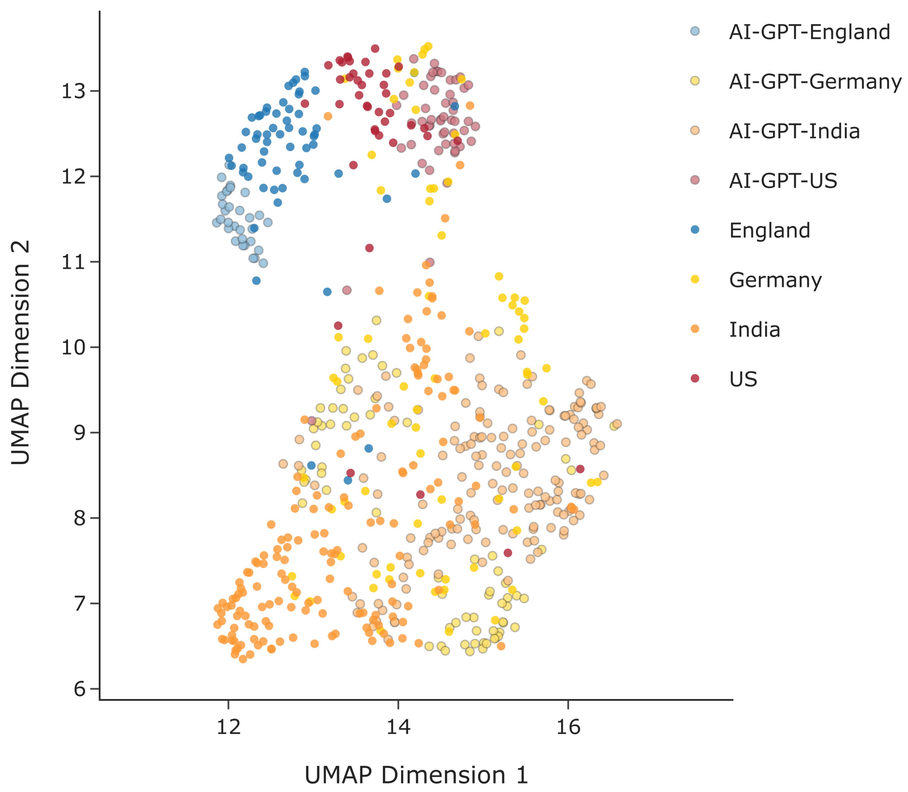
- "homogeneity-by-design" (Alvero et al. 2025: 60)
- possible homogenization of language in AI-modified and AI-generated texts, e.g. in favor of American English (Alvero et al. 2025, Mair 2024, Smith et al. 2025)
- emerging fears that
  - "non-dominant minor and emerging national standards" might not be able to "stand up to the powerful new anonymous standardising forces of AI-based machine translation and text-generation" (Mair 2024)
  - Generative AI technologies might "[reinforce] a hierarchy between language varieties and [affirm] "standard" varieties (particularly "standard" American English) as the most appropriate or legitimate form of communication" (Smith et al. 2025: 1)

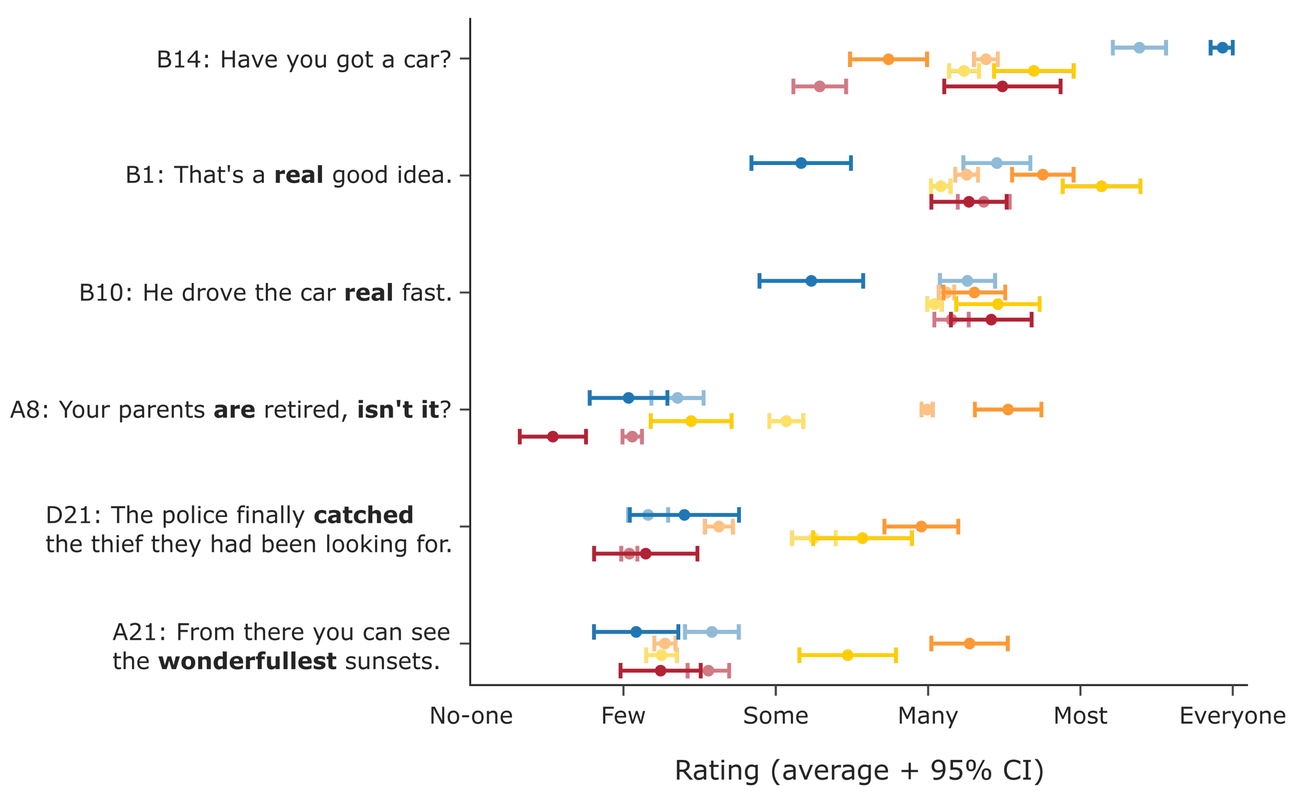
→ Are idiosynchrasies of varieties encoded in LLMs and can LLMs reproduce these when prompted?

Methods

- **Synthetic BSLVC datasets:** Generation of parallel data by prompting *ChatGPT* (model: gpt-5-mini; temperature: 1) to answer the BSLVC (Krug and Sell 2013) using real sociodemographic profiles of respondents from four varieties.
- **Variety-matched prompting:** For each BrE (31), AmE (52), Indian English (158), and German Learner English (72) case, the model received that respondent's age, gender, L1, parents' L1, qualification, and occupation and was tasked to answer according to their variety.
- **Dimensionality reduction:** Application of UMAP (McInnes et al. 2018) via the BSLVC dashboard (Vetter 2026) to compare synthetic and human BSLVC profiles (spoken section).
- **Spot-checks:** Inspection of ratings of selected variables such as control items or variables typical for a variety.

Results





Conclusion

- **Cluster structure:** Human and AI profiles cluster closely within each variety, forming adjacent, aligned, yet still distinguishable groupings.
- **Rating behaviour:** AI responses only partially reproduce variety-typical contrasts, are more conservative, less extreme (virtually no "no-ones"/"everyones"), and show markedly lower inter-participant variation.

Next steps

- **Identify distinguishing features:** Pinpoint linguistic features that reliably differentiate human and AI profiles.
- **Expand variety coverage:** Test additional English varieties included in the BSLVC database and include ChatGPT's default mode.
- **Evaluate prompting strategies:** Systematically test alternative prompting setups to assess their effect on AI–human alignment.

Funding

The BSLVC currently receives funding by the Deutsche Forschungsgemeinschaft (DFG) (Grant: 548274092).

&

BAKULE

Gefördert durch die Stiftung Innovation in der Hochschullehre

References

Alvero, Aj, Quentin Sedlacek, Maricela León, and Courtney Peña. 2025. "Digital accents, homogeneity-by-design, and the evolving social science of written language". Annual Review of Applied Linguistics 45: 50-68. DOI: 10.1017/S0267190525000042.

Krug, Manfred, and Katrin Sell. 2013. "Designing and Conducting Interviews and Questionnaires". In Research Methods in Language Variation and Change, edited by Manfred Krug and Julia Schlüter. Cambridge University Press. DOI: 10.1017/CBO9780511792519.007.

Mair, Christian. 2024. "ChatGPT for Linguists: Source of Data and Copilot for Analysis". Hermann Paul Centre for Linguistics, University of Freiburg.

McInnes, Leland, John Healy & James Melville (2018): UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv. doi:10.48550/ARXIV.1802.03426.

Smith, Genevieve, Eve Fleisig, Madeline Bossi, Ishita Rustagi, and Xavier Yin. 2025. "Standard Language Ideology in AI-Generated Language". arXiv:2406.08726. DOI: 10.48550/arXiv.2406.08726

Vetter, Fabian. 2026. BSLVC Dashboard. DOI: 10.17605/OSF.IO/4BUEF.